

## Bacterial Pleckstrin Homology Domains: A Prokaryotic Origin for the PH Domain

Qingping Xu<sup>1,2</sup>, Alex Bateman<sup>3</sup>, Robert D. Finn<sup>3</sup>, Polat Abdubek<sup>1,4</sup>, Tamara Astakhova<sup>1,5</sup>, Herbert L. Axelrod<sup>1,2</sup>, Constantina Bakolitsa<sup>1,6</sup>, Dennis Carlton<sup>1,7</sup>, Connie Chen<sup>1,4</sup>, Hsiu-Ju Chiu<sup>1,2</sup>, Michelle Chiu<sup>1,4</sup>, Thomas Clayton<sup>1,7</sup>, Debanu Das<sup>1,2</sup>, Marc C. Deller<sup>1,7</sup>, Lian Duan<sup>1,5</sup>, Kyle Ellrott<sup>1,5</sup>, Dustin Ernst<sup>1,4</sup>, Carol L. Farr<sup>1,7</sup>, Julie Feuerhelm<sup>1,4</sup>, Joanna C. Grant<sup>1,4</sup>, Anna Grzechnik<sup>1,7</sup>, Gye Won Han<sup>1,7</sup>, Lukasz Jaroszewski<sup>1,5,6</sup>, Kevin K. Jin<sup>1,2</sup>, Heath E. Klock<sup>1,4</sup>, Mark W. Knuth<sup>1,4</sup>, Piotr Kozbial<sup>1,6</sup>, S. Sri Krishna<sup>1,5,6</sup>, Abhinav Kumar<sup>1,2</sup>, David Marciano<sup>1,7</sup>, Daniel McMullan<sup>1,4</sup>, Mitchell D. Miller<sup>1,2</sup>, Andrew T. Morse<sup>1,5</sup>, Edward Nigoghossian<sup>1,4</sup>, Amanda Nopakun<sup>1,7</sup>, Linda Okach<sup>1,4</sup>, Christina Puckett<sup>1,4</sup>, Ron Reyes<sup>1,2</sup>, Christopher L. Rife<sup>1,2</sup>, Natasha Sefcovic<sup>1,6</sup>, Henry J. Tien<sup>1,7</sup>, Christine B. Trame<sup>1,2</sup>, Henry van den Bedem<sup>1,2</sup>, Dana Weekes<sup>1,6</sup>, Tiffany Wooten<sup>1,4</sup>, Keith O. Hodgson<sup>1,8</sup>, John Wooley<sup>1,5</sup>, Marc-André Elsliger<sup>1,7</sup>, Ashley M. Deacon<sup>1,2</sup>, Adam Godzik<sup>1,5,6</sup>, Scott A. Lesley<sup>1,4,7</sup> and Ian A. Wilson<sup>1,7\*</sup>

<sup>1</sup>Joint Center for Structural Genomics, <http://www.jcsg.org>

<sup>2</sup>Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

<sup>3</sup>The Wellcome Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

<sup>4</sup>Protein Sciences Department, Genomics Institute of the Novartis Research Foundation, San Diego, CA 92121, USA

Pleckstrin homology (PH) domains have been identified only in eukaryotic proteins to date. We have determined crystal structures for three members of an uncharacterized protein family (Pfam PF08000), which provide compelling evidence for the existence of PH-like domains in bacteria (PHb). The first two structures contain a single PHb domain that forms a dome-shaped, oligomeric ring with C<sub>5</sub> symmetry. The third structure has an additional helical hairpin attached at the C-terminus and forms a similar but much larger ring with C<sub>12</sub> symmetry. Thus, both molecular assemblies exhibit rare, higher-order, cyclic symmetry but preserve a similar arrangement of their PHb domains, which gives rise to a conserved hydrophilic surface at the intersection of the β-strands of adjacent protomers that likely mediates protein–protein interactions. As a result of these structures, additional families of PHb domains were identified, suggesting that PH domains are much more widespread than originally anticipated. Thus, rather than being a eukaryotic innovation, the PH domain superfamily appears to have existed before prokaryotes and eukaryotes diverged.

© 2009 Elsevier Ltd. Open access under [CC BY license](#).

\*Corresponding author. E-mail address: [wilson@scripps.edu](mailto:wilson@scripps.edu).

Abbreviations used: PH, Pleckstrin homology; PHb, bacterial PH domain; PTB, phosphotyrosine binding; VPS36, vacuolar protein sorting protein 36; DUF1696, domain of unknown function family 1696; JCSCG, Joint Center for Structural Genomics; MAD, multiwavelength anomalous diffraction; PEG, polyethylene glycol; asu, asymmetric unit; PDB, Protein Data Bank; PIPE, Polymerase Incomplete Primer Extension; TEV, tobacco etch virus; TCEP, tris(2-carboxyethyl)phosphine–HCl; SSRL, Stanford Synchrotron Radiation Lightsource; ALS, Advanced Light Source.

<sup>5</sup>Center for Research in  
Biological Systems, University  
of California, San Diego,  
La Jolla, CA 92093, USA

<sup>6</sup>Program on Bioinformatics and  
Systems Biology, Burnham  
Institute for Medical Research,  
La Jolla, CA 92037, USA

<sup>7</sup>Department of Molecular  
Biology, The Scripps Research  
Institute, La Jolla, CA 92037,  
USA

<sup>8</sup>Photon Science, SLAC  
National Accelerator  
Laboratory, Menlo Park,  
CA 94025, USA

Received 4 September 2009;  
received in revised form  
3 November 2009;  
accepted 4 November 2009  
Available online  
10 November 2009

**Edited by R. Huber**

**Keywords:** Pleckstrin homology (PH) domain; bacterial PH domain (PHb); higher-order symmetry; protein assembly; protein evolution

## Introduction

The Pleckstrin homology (PH) domain is a common protein module in eukaryotes found in proteins with a wide range of functions, involved in intracellular signaling and cytoskeletal organization.<sup>1,2</sup> The PH domain consists of a seven-stranded  $\beta$ -sandwich, which forms a pair of perpendicular  $\beta$ -sheets capped by a C-terminal amphipathic  $\alpha$ -helix.<sup>3–6</sup> PH domains are best known for binding phosphatidylinositol lipids and targeting proteins to the membrane.<sup>7</sup> The PH domain-like fold was also identified in proteins that had very low sequence similarities to conventional PH domains. According to the SCOP database,<sup>8</sup> they include phosphotyrosine binding (PTB) domain, FERM (band 4.1, ezrin, radixin, moesin homology) domain, Ran-binding domain, Enabled/VASP homology 1 domain, the GRAM domain of myotubularin, Dcp1 (decapping protein involved in mRNA degradation Dcp1), the GLUE domain of VPS36 (vacuolar protein sorting protein 36), the p62 subunit of TFIIF basal transcription factor complex, the POB3 subunit of the FACT complex, neurobeachin, and Necap1 (NECAP endocytosis associated 1). Although the overall topology of these PH-like domains is highly conserved, the loops connecting their core secondary structural elements are highly variable. PH-like domains, which define a superfamily of proteins with a PH domain fold (SCOP ID: 50729), are likely evolutionarily related to the PH domain. Proteins containing PH-like domains are essential in eukaryotes and PH-

like domains are among the most abundant modules (fourth) in the human genome.<sup>9</sup> As most characterized PH-like modules are involved in localizing proteins to specific cellular locations or protein partners, it was suggested that the PH-like module is a general targeting or “adaptor” domain.<sup>10</sup>

Despite the abundance of PH-like domains in eukaryotes, they had not been previously identified in bacteria. As part of our structural genomics effort in targeting novel protein families, we determined crystal structures of three bacterial proteins from a previously structurally uncharacterized protein family Pfam PF08000 [also known as domain of unknown function family 1696 (DUF1696)],<sup>11</sup> which appears to be involved in the bacterial cell envelope stress response. These structures share a common domain that is surprisingly similar to the eukaryotic PH-like domains, thus providing the first direct evidence for the existence of PH-like domains in prokaryotes. We, therefore, suggest a much older origin for PH-like domains by illustrating that bacterial PH-like (PHb) domains are likely evolutionarily related to eukaryotic PH-like domains. The PHb proteins form oligomeric ring assemblies of different sizes and symmetries, providing an excellent example of how molecular architecture can evolve, where the general form of the molecular assembly is preserved, but its size is expandable. Furthermore, we show that several other uncharacterized protein families are likely to contain PHb domains. With more than 1000 prokaryotic homologs identified so far, it seems that this structural module is also widespread among bacteria and archaea.

## Results

### Sequence analysis of the bacterial PF08000 family

PF08000 defines a small protein family whose members are distributed in bacteria, bacteriophages, and archaea, with the majority in firmicutes and proteobacteria. As this family is of unknown function, as well as being structurally uncharacterized, it is also classified as DUF1696 in the Pfam database. One homolog has been identified in *Methanobrevibacter smithii*, an archaeal member of the human gut microbiota. Most members of this family have a sequence length of ~125 residues (group 1), but a second group is longer (~204 residues) with a C-terminal extension, and such sequences are detected only in Bacillaceae (e.g., *Bacillus subtilis*). Similarities in the core residues suggest that both groups share a common evolutionary origin, but group 2 likely has evolved more recently due to its narrower distribution in bacteria and its more restricted sequence diversity (~50% identical). A third group of more distant homologs, containing the core sequence of the first group at their C-termini, is present in phages, such as *Lactococcus* phage bIL285 (ORF2, 181 residues) and *Staphylococcus* phage X2 (ORF15, 241 residues).

### Structure determination and quality of the models

To gain insights into this family of proteins, we determined crystal structures of three bacterial members (that represent groups 1 and 2) of the PF08000 family at the Joint Center for Structural Genomics (JCSG†) using its high-throughput structural genomics pipeline.<sup>12</sup> The proteins from *Shewanella loihica* (*sl*), *Shewanella amazonensis* (*sa*), and *Exiguobacterium sibiricum* (*es*) were expressed in *Escherichia coli* as selenomethionine derivatives. As full-length (*sl*) and (*es*) failed to crystallize, different constructs with serial truncations of up to 16 residues at the N- and/or C-terminus, in steps of 4 residues, were screened in parallel. Only one N-terminal truncation construct of each protein, residues 9–124 for (*sl*) and 13–204 for (*es*), resulted in successful structure determination. Structure determination of full-length (*sa*) was enabled using reductive methylation of the protein prior to crystallization. All three structures were solved independently using the multiwavelength anomalous diffraction (MAD) phasing method.

To our surprise, these structures had similarity to eukaryotic PH-like domains that we now define as bacterial PH domain-like proteins (Fig. 1) from group 1 (PHb1) [*S. loihica* [PHb1(*sl*), residues 9–124] and *S. amazonensis* [PHb1(*sa*), residues 1–125]] and from group 2 (PHb2) [*E. sibiricum* 255–15 (PHb2,

residues 13–204)]. The two PHb1s are closely related (80% sequence identity); PHb1(*sa*) can be aligned to PHb1(*sl*) without gaps except for an extra glycine at its C-terminus (Fig. 2).

Both PHb1 structures were solved in orthorhombic space group  $P2_12_12_1$  at 2.0 Å resolution with one pentamer in the asymmetric unit (asu) [ $R_{\text{cryst}} = 18.6/R_{\text{free}} = 22.7$  for PHb1(*sl*), Protein Data Bank (PDB) ID: 3dcx;  $R_{\text{cryst}} = 19.1/R_{\text{free}} = 23.3$  for PHb1(*sa*), PDB ID: 3hsa]. PHb2 (PDB ID: 3b77) was solved in tetragonal space group  $P4$  at 2.42 Å resolution with six molecules per asu (two 1/4 dodecamers) ( $R_{\text{cryst}} = 21.6/R_{\text{free}} = 25.4$ ). The mean residual errors of the coordinates for PHb1(*sl*), PHb1(*sa*), and PHb2 were estimated at 0.16, 0.17, and 0.24 Å, respectively, by an  $R_{\text{free}}$ -based diffraction-component precision index method.<sup>13</sup> Analysis of the resulting models with MolProbity<sup>14</sup> indicated good geometry with an all-atom clash score of 6.90, 0.4% rotamer outliers, and 98.0% of residues (no outliers) in the favored region of the Ramachandran plot for PHb1(*sl*) [NB: similar scores for PHb1(*sa*)] and an all-atom clash score of 6.82, 1.9% rotamer outliers, and 97.7% of residues in the favored region of the Ramachandran plot (five outliers in regions of poor density) for PHb2. The final model of PHb1(*sl*) contains 561 amino acid residues, 1 chloride ion, 7 (4S)-2-methyl-2,4-pentanedioles, and 322 waters in the asu. The PHb1(*sa*) model consists of 585 protein residues, 2 polyethylene glycol (PEG) fragments, 6 glycerols, and 242 waters, and PHb2 has 1106 residues and 173 waters. Data collection, model, and refinement statistics are summarized in Table 1 (and Tables S1–S3).

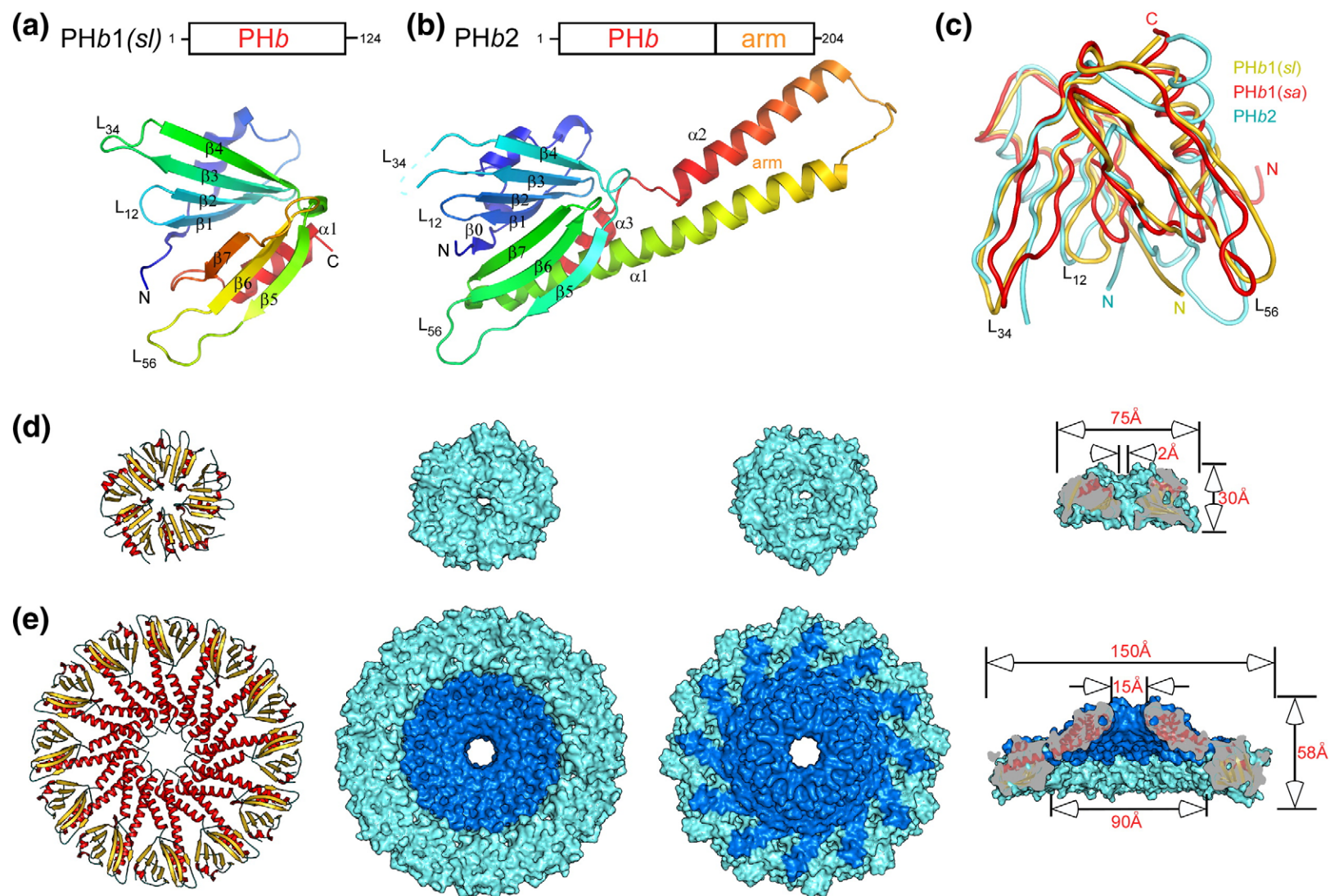
### PHb1 and PHb2 monomers

PHb1 and PHb2 share a common domain that is structurally similar to eukaryotic PH domains (Fig. 1a–c), which we have defined as a PHb domain (PH domain in bacteria). PHb1(*sl*) (residues 9–124) has only a single PHb domain that consists of the canonical seven  $\beta$ -strands in two  $\beta$ -sheets ( $\beta_1$ – $\beta_4$  and  $\beta_5$ – $\beta_7$ ) and a C-terminal helix ( $\alpha_1$ ) (Fig. 1a). It has an additional short  $\beta$ -strand  $\beta_0$  that connects to  $\beta_1$  with a large loop containing helix  $\alpha_0$  that packs against the  $\beta_1$ – $\beta_4$  surface. PHb2 contains the PHb domain at its N-terminus (residue 14–136) and a helical-hairpin structure (~68 residues) at the C-terminus (Fig. 1b) that extends out and returns as two  $\alpha$ -helices ( $\alpha_1$ ,  $\alpha_2$ ) to the PHb domain. A short helix at the C-terminus ( $\alpha_3$ ) docks between  $\alpha_1$  and the  $\beta_2$ – $\beta_3$  ( $L_{23}$ ) loop and, as a result, links the other end of the hairpin to the PHb domain. This hairpin attachment is crucial for formation of the PHb2 homododecamer.

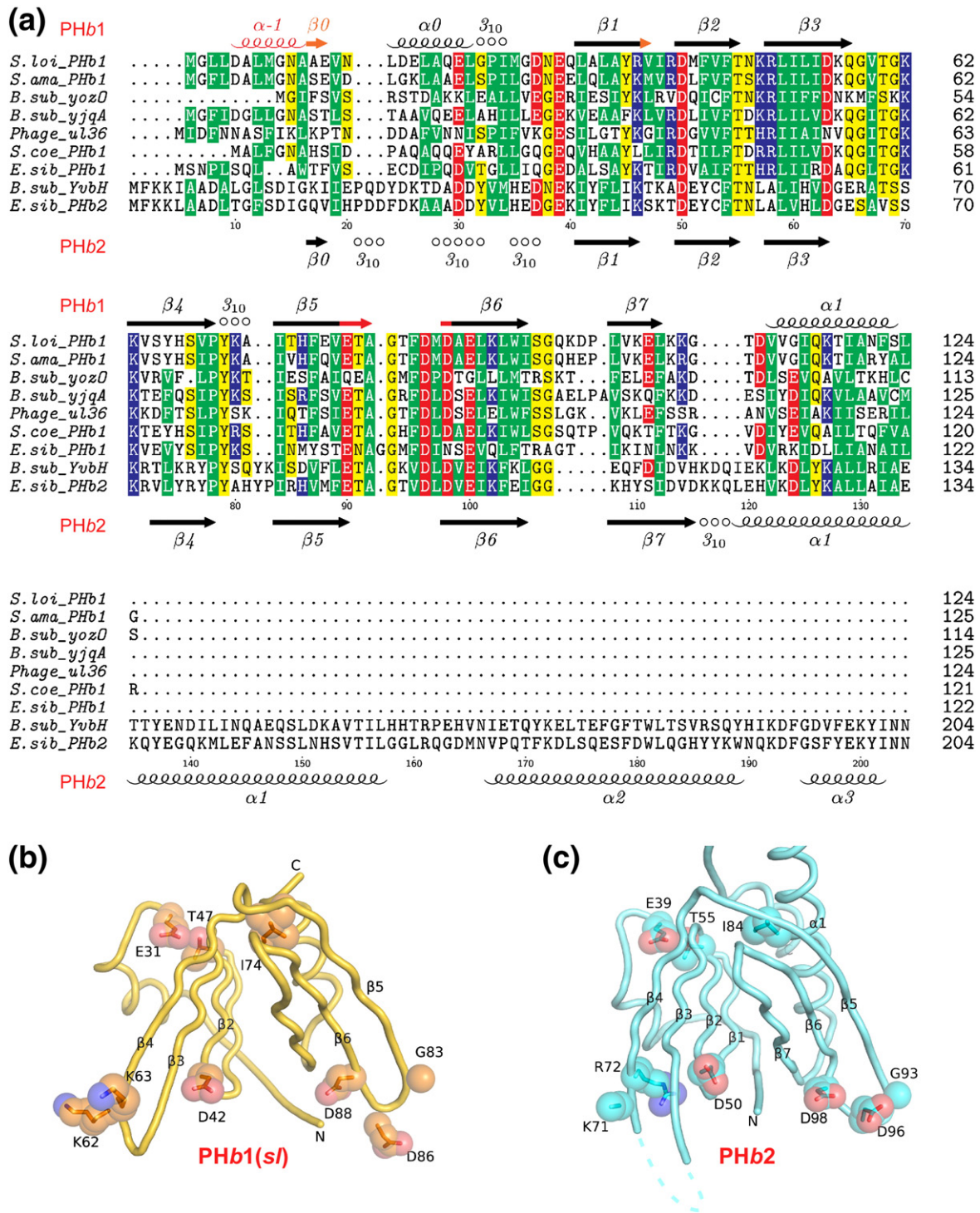
The overall shape of the PHb domain resembles a butterfly of two  $\beta$ -sheets with the  $\beta_3$ – $\beta_4$  loop ( $L_{34}$ ) and the  $\beta_5$ – $\beta_6$  loop ( $L_{56}$ ) at the tip of each wing. The PHb domains in two PHb1 orthologs contain a structurally conserved core, except for the N-termini (before  $\alpha_0$ ), the  $L_{34}$  loop, and the  $L_{56}$  loop regions that are highly variable in conformation (Fig. S1).

† <http://www.jcsg.org>





**Fig. 1.** Structures of PHb1s and PHb2. (a) Ribbon diagram of PHb1(*sl*) monomer color coded from N-terminus (blue) to C-terminus (red). A schematic diagram of the domain architecture of PHb1(*sl*) is shown at the top. (b) Ribbon diagram of PHb2 monomer. (c) A structural superposition of the PHb domains of PHb1(*sl*) (gold), PHb1(*sa*) (red), and PHb2 (cyan). (d) PHb1(*sl*) pentamer shown in ribbons (bottom or base view) and surface representations (bottom, top and side views). (e) PHb2 dodecamer shown in same representation as PHb1. The PHb domain is colored cyan, as in PHb1. The PHb2 helical-hairpin arm attachment (residues 136–204) is colored blue.



**Fig. 2.** Sequence conservation of PHb homologs. (a) Sequence alignment of PHbs from *S. loihica*, *S. amazonensis*, *B. subtilis* (YozO, YjqA, and YvbH), *Lactococcus phage ul36.k1*, *Streptomyces coelicolor*, and *E. sibiricum*. The secondary-structure elements of the two PHb1s are shown at the top [red, unique to PHb1(sa); orange, unique to PHb1(sl); black, common]; secondary structure and sequence numbering of PHb2 are shown at the bottom. The conserved residues are highlighted (green, hydrophobic; yellow, polar; blue, basic; red, acidic). (b) Mapping of highly conserved residues onto the PHb domain of PHb1 with the side chains shown in ball and stick surrounded by their van der Waals surface. (c) Mapping of highly conserved residues onto the PHb domain of PHb2, color coded as in (b).

The 10 copies of the PHb1 monomers [from the two pentamers of PHb1(sl) and PHb1(sa)] can be superimposed with an average rmsd of 1.31 Å for 94

aligned C $\alpha$  atoms; the rmsd values between different monomers are significantly higher if all variable regions were included (Tables S4–S6). The N-

**Table 1.** Summary of data collection and refinement statistics

	<i>S. loihica</i> PHb1	<i>S. amazonensis</i> PHb1	<i>E. sibiricum</i> PHb2
Structure (PDB ID)	(3dcx)	(3hsa)	(3b77)
Beamline	SSRL BL 11-1	SSRL 9-2	ALS 8.2.2
Space group	$P2_12_12_1$	$P2_12_12_1$	$P4$
Unit cell parameters (Å)			
<i>a</i>	61.07	33.23	150.99
<i>b</i>	75.32	129.49	150.99
<i>c</i>	139.40	138.73	76.20
<i>Data collection</i>	3dcx-infl	3hsa-peak	3b77-infl
Wavelength (Å)	0.9793	0.9792	0.9799
Resolution range (Å)	29.83–2.00	47.40–2.00	47.73–2.42
No. of observations	177,704	150,785	245,189
No. of unique reflections	43,831	41,702	65,459
Completeness (%) <sup>a</sup>	99.1 (97.9)	98.9 (98.6)	99.7 (99.8)
Mean $I/\sigma(I)$ <sup>a</sup>	13.9 (2.3)	12.1 (2.6)	10.6 (1.9)
$R_{\text{sym}}$ on $I$ (%) <sup>a</sup>	6.7 (68.4)	8.8 (49.3)	8.1 (76.1)
Highest-resolution shell	2.11–2.00	2.10–2.00	2.55–2.42
<i>Model and refinement statistics</i>			
Resolution range (Å)	29.83–2.00	47.4–2.0	47.73–2.42
Cutoff criteria	$ F  > 0$	$ F  > 0$	$ F  > 0$
No. of reflections (total)	43,782	41,641	65,460
No. of reflections (test)	2200	2115	3326
Completeness (% total)	98.9	98.5	99.7
$R_{\text{cryst}}$	18.4	19.1	21.4
$R_{\text{free}}$	22.7	23.6	25.4
<i>Stereochemical parameters</i>			
Restraints (RMS observed)			
Bond length (Å)	0.014	0.015	0.015
Bond angle (°)	1.41	1.66	1.50
Average isotropic $B$ -value <sup>b</sup> (Å <sup>2</sup> )	35.4	27.5	65.6
ESU based on $R_{\text{free}}$ (Å)	0.16	0.17	0.24
Chains/protein residues/atoms	5/561/4481	5/585/4675	6/1106/8993
Solvent molecules	330	250	173

ESU, estimated overall coordinate error.

$$R_{\text{sym}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$$

$R_{\text{cryst}} = \frac{\sum_{hkl} | |F_{\text{obs}}| - |F_{\text{calc}}| |}{\sum_{hkl} |F_{\text{obs}}|}$ , where  $F_{\text{calc}}$  and  $F_{\text{obs}}$  are the calculated and observed structure factor amplitudes, respectively.

$R_{\text{free}}$  = as for  $R_{\text{cryst}}$ , but for 5.0% of the total reflections chosen at random and omitted from refinement.

<sup>a</sup> Statistics for the highest-resolution shell are in parentheses.

<sup>b</sup> This value represents the total  $B$  that includes TLS and residual  $B$  components.

terminal residues, which were deleted in the PHb1 (*sl*) construct, are ordered in two monomers (A and B) of PHb1(*sa*) due to crystal contacts and form a short helix ( $\alpha$ -1, residues 5–11). The conformation of the N-terminal region differs significantly between PHb1(*sa*) and PHb1(*sl*) (Fig. 1c). PHb1(*sa*) does not contain the first strand ( $\beta$ 0) found in PHb1(*sl*). These structural variations, which likely reflect differences in the contact environments in the crystals, demonstrate the inherent flexibility of the PHb domains (Fig. S2). On the other hand, the six independent protomers in the asu of PHb2 have similar packing environment and are essentially identical with each other in structure with an average rmsd of 0.27 Å for 180 aligned C $\alpha$  atoms. The ordered L<sub>56</sub> loops in PHb2 are also involved in crystal contacts, although to a lesser extent than PHb1(*sl*). The L<sub>34</sub> loop of PHb2 is exposed to solvent and disordered. The PHb domains of PHb1 and PHb2 can be superimposed with an average rmsd of ~3.0 Å for ~110 aligned C $\alpha$  atoms (Table S7). The main differences between the two PHb domains in PHb1(*sl*) and PHb2 are found in  $\alpha$ 0, the second  $\beta$ -sheet ( $\beta$ 5– $\beta$ 7),  $\alpha$ 1, and the loop regions (Fig. 1c).

Mapping of conserved residues within the PHb family onto the structures of PHb1 and PHb2 shows high conservation of residues in strand  $\beta$ 2, the L<sub>34</sub> loop, and the L<sub>56</sub> loop (Fig. 2). A highly conserved glutamate (Glu31 in PHb1 and Glu39 in PHb2) is buried inside the protein where its carboxyl side chain forms multiple hydrogen bonds with the main-chain amide (NH) of a conserved threonine (Thr47 in PHb1 and Thr55 in PHb2) and the side chain of an arginine (Arg50 in PHb1 and Arg76 in PHb2). These interactions, combined with the conserved hydrophobic core, are expected to be important for the structural integrity of PHb domains. Furthermore, Arg50 and Arg76 also participate in a water-mediated hydrogen-bond network at the oligomeric interfaces. The conserved residues that are clustered near the L<sub>34</sub> and L<sub>56</sub> loops are solvent exposed and are, thus, more likely to be of functional significance.

### Comparisons of PHb with eukaryotic PH-like domains

DALI structural similarity searches,<sup>15</sup> limited to the PHb domain from either PHb1 or PHb2, iden-

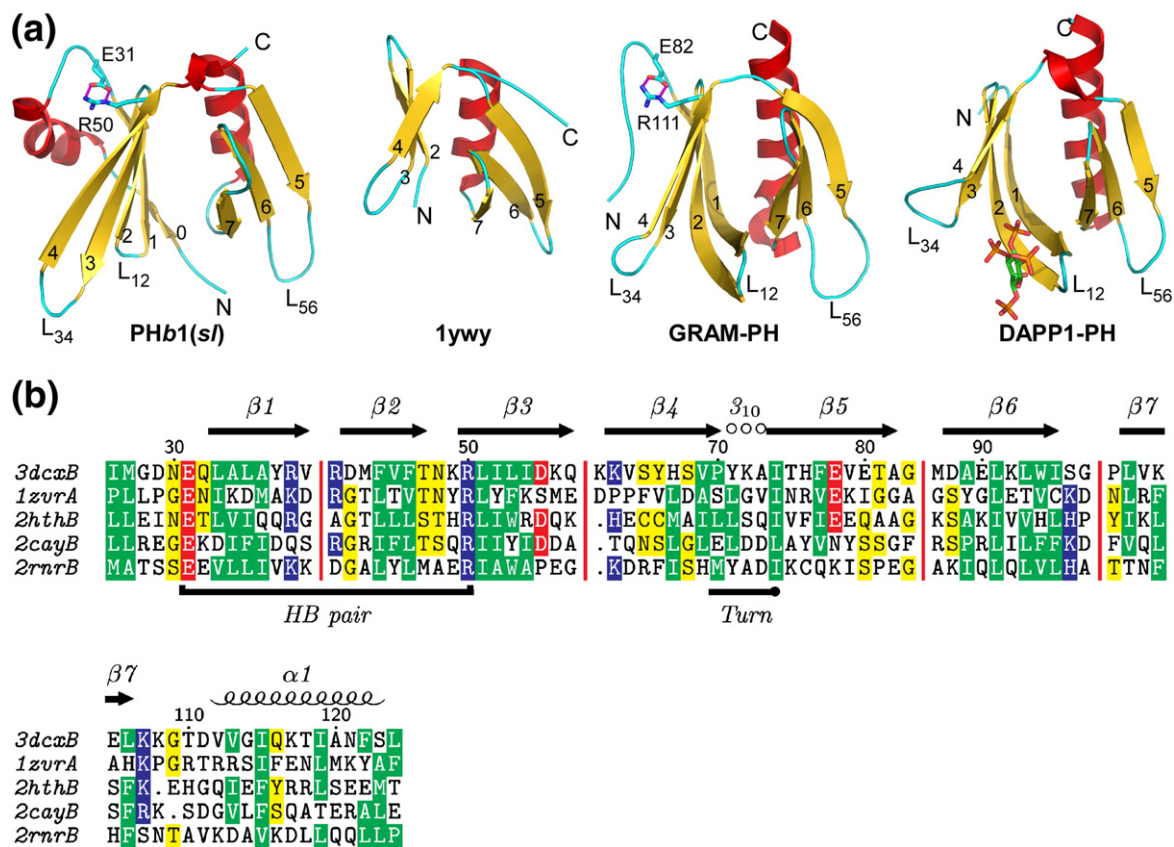


tified a large number of proteins from the eukaryotic PH superfamily. The top hits included proteins of diverse function, such as the GRAM-PH domain of myotubularin family phosphoinositide phosphatase MTMR2<sup>16</sup> (PDB ID: 1zvr,  $Z=9.9$ ,  $\text{rmsd}=2.3$  Å for 94 aligned C $^{\alpha}$  atoms, 14% sequence identity), GLUE domains of VPS36 of human and *Saccharomyces cerevisiae*<sup>17,18</sup> (PDB ID: 2hth,  $Z=9.6$ ,  $\text{rmsd}=2.1$  Å for 90 aligned C $^{\alpha}$  atoms, 16% id; PDB ID: 2cay,  $Z=9.3$ ,  $\text{rmsd}=2.4$  Å for 90 aligned C $^{\alpha}$  atoms, 13% id), and Dcp1<sup>19</sup> (PDB ID: 2qkl,  $Z=8.8$ ,  $\text{rsmd}=2.3$  Å for 93 aligned C $^{\alpha}$  atoms). Three previously solved structures of uncharacterized bacterial proteins, two by the JCSG (PDB IDs: 2ra9 and 2re3) and one by the Northeast Structural Genomics Consortium (PA2021, PDB ID: 1ywy), also appear to contain a domain with PH fold but have one less  $\beta$ -strand ( $\beta$ 1) and are structurally more distant. The closest eukaryotic protein in terms of structural similarity to PA2021 is the Ras-related protein RalA<sup>20</sup> (PDB ID: 1zc3,  $Z=3.4$ ,  $\text{rmsd}=2.5$  Å for 59 aligned C $^{\alpha}$  atoms, 10% id). Structural comparisons of the PHb1 domain with PA2021 and representative eukaryotic PH-like domains are shown in Fig. 3. The conserved core consists of 7  $\beta$ -strands ( $\beta$ 1– $\beta$ 7) and a C-terminal  $\alpha$ -helix ( $\alpha$ 1). Like eukaryotic PH-like domains, the

$\beta$ 1– $\beta$ 4 sheet of PHbs packs perpendicularly with the  $\beta$ 5– $\beta$ 7 sheet through hydrophobic contacts. An additional connection between the two  $\beta$ -sheets is mediated through the conserved C-terminal  $\alpha$ -helix ( $\alpha$ 1). The most significant structural difference between PHb domains and canonical eukaryotic PH-like domains is the length and conformation of the  $\beta$ 1– $\beta$ 2 hairpin. The longer hairpin of the eukaryotic PH-like domain curves towards the open edge of the  $\beta$ 5– $\beta$ 7 sheet, such that  $\beta$ 1 interacts with  $\beta$ 7 to form an extended antiparallel  $\beta$ -sheet, thus producing a closed barrel. In contrast, the  $\beta$ 1– $\beta$ 2 hairpin of PHbs, including the two-residue L<sub>12</sub> loop, is shorter and does not interact with the  $\beta$ 7 strand directly. The first ~30 N-terminal residues of PHbs are not conserved across the PH-like superfamily of proteins. However, some proteins in the PH-like family do contain extra structural elements that are functionally equivalent for stabilization of the  $\beta$ 1– $\beta$ 4 surface. For example, Dcp1 contains a helix that is equivalent to  $\alpha$ 0 of PHb1.

### Oligomeric rings of PHb

PHb1 and PHb2 form oligomeric rings with rarer, higher-order, cyclic symmetry.<sup>21</sup> The structures and



**Fig. 3.** Comparisons of bacteria PH domains with eukaryotic PH domains. (a) Representative PH domains shown in same orientations: PHb1(sl) (PDB ID: 3dcx), uncharacterized protein PA2021 from *Pseudomonas aeruginosa* (PDB ID: 1ywy), the GRAM-PH domain myotubularin of (PDB ID: 1zvr), and the PH domain of DAPP1/PHISH complexed with inositol-(1,3,4,5)-tetrakisphosphate (PDB ID: 1fao). (b) Structure-based sequence alignment of PHb domain (PDB ID: 3dcx) with top hits of a DALI search: GLUE-PH domains (PDB ID: 2cay and 2hth), GRAM-PH (PDB ID: 1zvr), and TFIIF-PH (PDB ID: 2rnr). Red vertical bars indicate omitted gaps in the alignment.

dimensions of the PHb1(*sl*) pentamer and the PHb2 dodecamer are illustrated in Fig. 1d and e. PHb1(*sl*) and PHb1(*sa*) form a disk with non-crystallographic C<sub>5</sub> symmetry that is normally only encountered in viruses with icosahedral capsids. A hole in the center of the pentamer narrows from the top towards the base. The PHb1(*sl*) pentamer buries a total surface area of 9080 Å<sup>2</sup> (34% of total surface area, 1816 Å<sup>2</sup> per monomer). The PHb1(*sa*) pentamer is similar to that of PHb(*sl*); thus, we refer only to PHb1(*sl*) in the discussion hereafter. PHb2 is also a dome-shaped assembly, but with C<sub>12</sub> cyclic symmetry that is generated by crystallographic C<sub>4</sub> axis and a non-crystallographic C<sub>3</sub> axis. The PHb2 assembly is significantly larger with an outer diameter of 150 Å, that is, twice that of PHb1(*sl*). The PHb2 dodecamer buries a total surface area of 46,700 Å<sup>2</sup> or ~50% of total surface area (3890 Å<sup>2</sup> per monomer). Its central channel is also significantly larger with a diameter of 90 Å at the base and 15 Å at the top (NB: the actual opening would likely be smaller if the Arg161 side chains around the top channel were ordered). The openings of both central channels are guarded by positively charged residues [Lys72 and Lys99 in PHb1(*sl*) and Arg161 in PHb2]. Both oligomeric states are consistent with size-exclusion chromatography experiments. Despite significant differences in the sizes of the two assemblies, the orientations of the PHb domains within the rings are highly similar. The base of each ring consists of repeating β-blades, while the top of each ring is primarily helical. If we superpose one subunit from the PHb1(*sl*) pentamer with one from the PHb2 dodecamer, then the average rotation angle that relates adjacent monomers from the two assemblies is 43° (standard deviation of 1.6°; range, 41–45°), which is very close to the expected value of 42° (360°/5–360°/12), if the deviation in rotation angle between monomers is uniform over the pentamer.

The oligomerization mode of the PHb domain resembles those of other intermolecular β-sheet forming assemblies of structurally different folds, for example, the Sm-like ribonucleoproteins (SCOP ID: 50182) and the TRAP-like superfamily of proteins (SCOP ID: 51219) in SCOP.<sup>8</sup> These propeller-like assemblies also possess high-order cyclic symmetries and show variations of the number of subunits in the ring. GroES also has a dome-shaped architecture;<sup>22</sup> however, the fold of the GroES monomer and the assembly details of the heptamer are different from those of PHbs.

### PHb1 pentamer

The formation of the pentamer in PHb1(*sl*) is mediated primarily through hydrogen-bonding interactions (Fig. 4a and b). β4 and β5 of two adjacent protomers form an antiparallel β-sheet, consisting of five hydrogen bonds between their main-chain atoms. As a result, the β0–β4 and β5–β7 of the next protomer form an extended eight-stranded β-sheet, arranged like the blades of a propeller (β-blade) at the base of the pentamer.

Additionally, His67 hydrogen bonds with His76 on the adjacent strand. The Thr81 hydroxyl group interacts with a main-chain amide. This conserved interaction (Thr91 in PHb2) helps to stabilize the β4 N-terminus with the β5 C-terminus of the next protomer and is likely important for PHb function since the proposed binding site is in close proximity (see below).

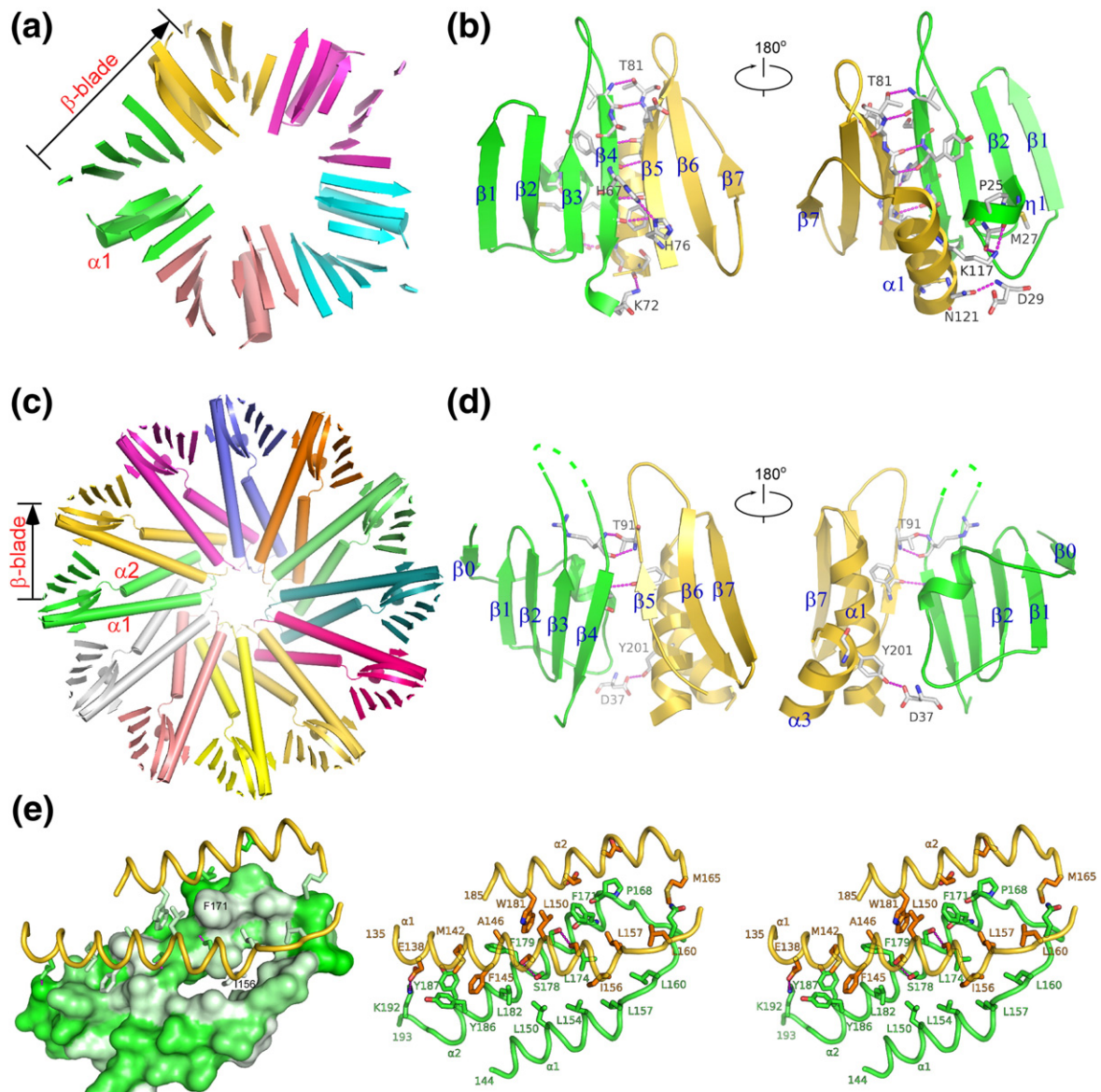
The second contact point in the pentamer involves α1 (residues 113–121) from one monomer and the 3<sub>10</sub> helix region (residues 25–29) of the adjacent protomer. Lys117 adopts different conformations in different promoters and makes polar contact with different partners, such as the carbonyl groups of Met27 and Gly24 and the side chain of Asp29. Asn121Oδ1 forms another hydrogen bond with the main-chain amide of Asp29. Additionally, a water-mediated hydrogen-bonding network is observed for polar/charged residues buried in the interface (Arg50, Asn116, and Ser68). Overall, the PHb1(*sl*) pentamer interfaces lack specific, conserved, side-chain interactions, except for Thr81.

### Ring expansion in PHb2 through addition of a helical hairpin

The PHb2 α1 is significantly longer and forms a hairpin appendage with an additional helix α2 at the C-terminus (distance=6.8 Å and Ω=–159° to α1). The hairpin constitutes the primary mediator of the central core of the dodecameric PHb2 ring, mainly through hydrophobic interactions (Fig. 4c and e). Interactions between the α1 and α2 hairpins contribute to 63% (29,350 Å<sup>2</sup>) of the overall buried surface within the dodecamer. The packing of hairpins involves the docking of surface hydrophobic residues of α1 (around Ile156) and α2 (Val167, Phe171, and Phe179) into a hydrophobic pocket formed by the adjacent hairpin of the neighboring protomer (Fig. 4e). Helix α2 (near Ser175–Ser178) intersects with α1 of the next protomer (near Ser149–Ser153), with an inter-helix distance of 7.9 Å and an angle (Ω) of 134°. The four serines above are strategically located since their small side chain permits the helices to pack closely together and also form hydrogen bonds with each other, thus further stabilizing the interface.

Despite significant difference in the size of the oligomers, the arrangement of the PHb domains in the PHb2 dodecamer is remarkably similar to that of the PHb1 pentamer (Fig. 4c and d). The intersubunit eight-stranded β-sheets, located on the periphery of the ring, are arranged in a similar manner with respect to β4 and β5 of the two neighboring protomers. However, only two main-chain hydrogen bonds [compared to five in PHb1(*sl*)] are formed between the two strands due to the separation at the N-terminus of β4 and C-terminus of β5 to accommodate the packing of the α1–α2 hairpins. The second interface between PHb domains in PHb2 also resembles that of the PHb1 and involves contacts between the N-terminal 3<sub>10</sub> helix of PHb (residues 31–37) and a section of helix α1 (residues 120–130) of





**Fig. 4.** Dissection of PHb assembly interfaces. (a) Ribbon diagram showing the  $\beta$ -sheets and  $\alpha 1$  of the PHb1(s) pentamer. (b) Interface between two adjacent PHb1(s) protomers (green and gold). Hydrogen bonds are shown as dashes (magenta). (c) Ribbon diagram showing the  $\beta$ -sheets and  $\alpha 1$ – $\alpha 2$  of the PHb2 dodecamer. (d) Interface between two adjacent PHb2 protomers (green and gold). (e) Stereo view of interactions between  $\alpha 1$  and  $\alpha 2$  hairpins of two adjacent protomers (right). A combined surface/ribbon representation of the same view is shown on the left. The surface is colored according to hydrophobicity, where greenish gray is the most hydrophobic.

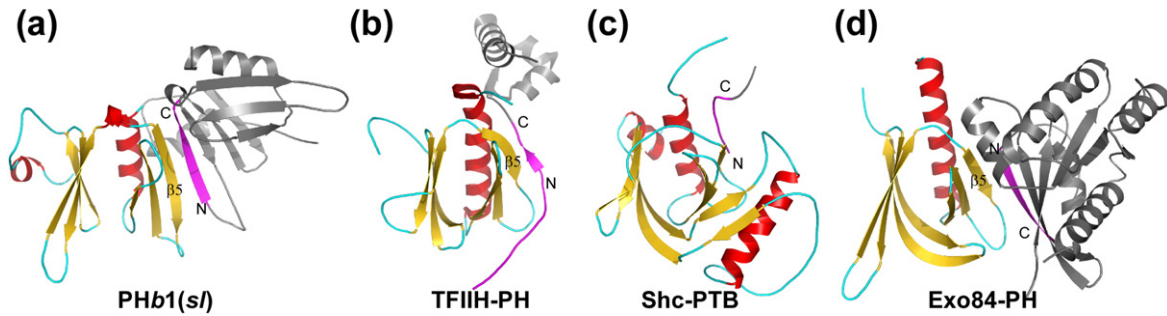
two adjacent promoters. The interface involves van der Waals contacts (33% of nonpolar residues) but no hydrogen bonds. A region of helix  $\alpha 3$  (residues 200–201), which is unique to PHb2, also contributes to these interactions (Tyr201/Asp37).

An additional structural adaptation by the PHb domain of PHb2 was observed, presumably to facilitate formation of the dodecamer ring (Fig. 1c). The retraction of the  $\beta 3$ – $\beta 4$  and  $\beta 5$ – $\beta 6$  hairpin wings and the N-terminal inter-protomer contact region (residues 31–37) towards the center reduces the overall width of the PHb domain. A small change in the orientation of  $\alpha 1$  towards the C-terminus widens the gap between  $\alpha 1$  and L<sub>23</sub> and allows docking of  $\alpha 3$ . PHb2 thus represents a

fascinating example of protein evolution by which a new molecular architecture is derived from a simple modification of a basic module.

#### Similarity of PHb oligomerization to protein recognition by eukaryotic PHs

The oligomerization of PHbs involves an extension of the  $\beta$ -sheets at the open edge of  $\beta 4$  or  $\beta 5$ , with additional contributions from the C-terminal  $\alpha$ -helix ( $\alpha 1$ ) and an N-terminal region containing a  $3_{10}$  helix. Interestingly, this mode of protein–protein interaction, involving the open edge of the  $\beta 5$  strand, is commonly observed in eukaryotic PH-like domains (Fig. 5). In PTB domains, peptide ligands are bound



**Fig. 5.** The oligomerization interface of PHb overlaps with a common protein interaction site of eukaryotic PH domains. The superimposed PH domains are shown in yellow/red/cyan. The protein/peptide partners are shown in gray with the interface elements highlighted in magenta. (a) PHb1(*sl*) dimer (PDB ID: 3dcx). (b) TFIIH (PH) complexed with TFIIIE (PDB ID: 2mr). (c) Shc PTB domain complexed with a phosphor-peptide (PDB ID: 1shc). (d) Exo84 PH domain complexed with RalA (PDB ID: 1zc3).

as an antiparallel, pseudo- $\beta$ -sheet with extensive contacts with the  $\beta 5$  strand and the C-terminal  $\alpha$ -helix.<sup>23</sup> Radixin utilizes the same interface, such that a  $\beta$ -strand binds the shallow groove between  $\beta 5$  and  $\alpha 1$ .<sup>24,25</sup> Corresponding to the C-terminal  $\beta$ -turn NPxY motif of the PTB peptides, a conserved structural motif (<sup>70</sup>PYxxI<sup>74</sup>) exists at the C-terminus of  $\beta 4$  in PHb1 and adopts a similar main-chain conformation to PTB peptides, such as in Shc.<sup>26</sup> However, the role of the NPxY motif tyrosine in PTB and PHbs is different; the tyrosine in the PTB peptides plays a significant role in PTB binding, while the tyrosine is buried inside PHbs and not involved in protein-protein interaction.

Another example of the  $\beta 5$ - $\beta 7$   $\beta$ -sheet extension at  $\beta 5$  is observed in the recent structure of the TFIIH p62 subunit (PH domain) in complex with the C-terminal acidic domain of the general transcription factor TFIIIE.<sup>27</sup> Other modes of utilization of the  $\beta 5$  edge for protein-protein interaction were also observed. For example, Ral-binding domain of Exo84 utilizes the  $\beta 5$  edge for Ral interaction with the peptide from Ral adopting a parallel conformation.<sup>20</sup> The interaction of the GLUE domain of VPS36 with ubiquitin does not directly involve a  $\beta$ -sheet type of interaction, but ubiquitin also occupies the  $\beta 5$  edge of the GLUE domain.<sup>17</sup> Thus, the PHb oligomerization interface overlaps with a common site of protein-protein interaction in eukaryotic PH domains. Furthermore, the mode of PHb oligomerization through  $\beta$ -sheet extension is similar to the mode of protein recognition by eukaryotic PH domains.

### A conserved binding site

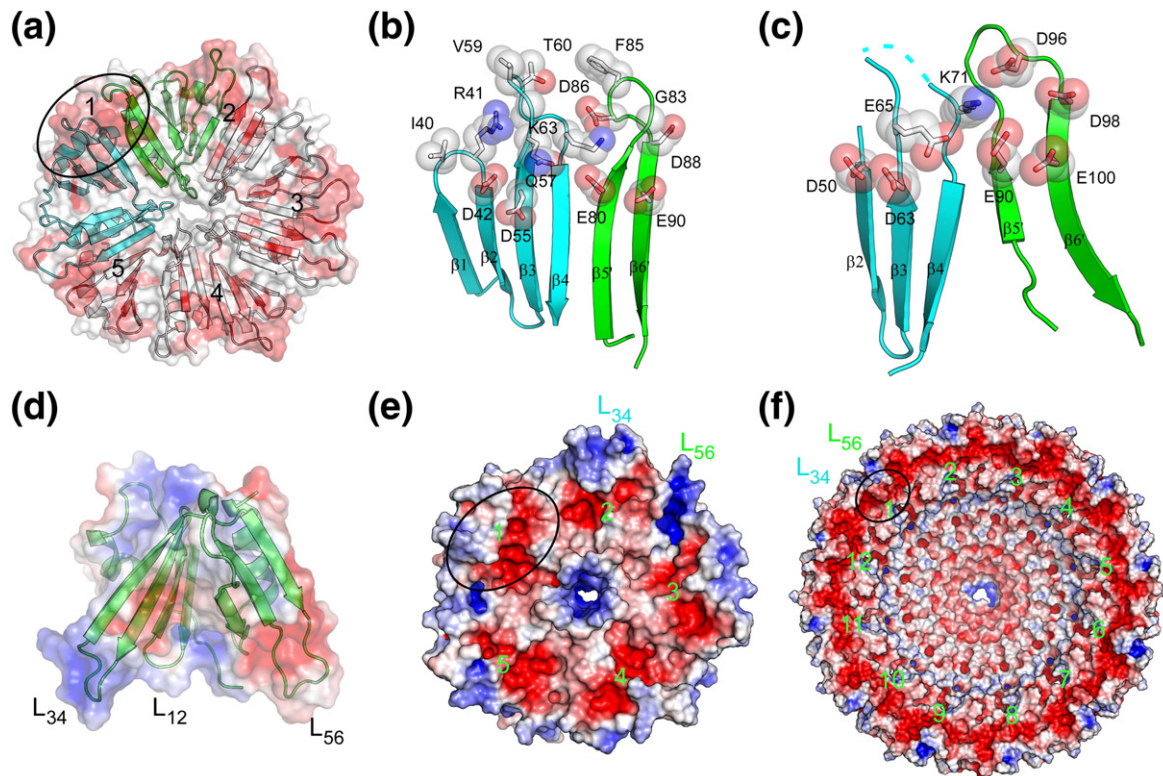
Residues that are important for the structural integrity and function of a protein family are often under strong evolutionary constraints. Thus, we should be able to predict the functional sites of PHb assemblies by analyzing the sequence conservation patterns in the context of the protein structure, particularly for clustering of conserved residues on the protein surface. Aside from the few buried, charged residues discussed above, all conserved surface residues cluster to a single location on each

$\beta$ -blade at the base of the dome (Fig. 6a). Each site is formed by a contribution from two adjacent protomers, thus clearly indicating the physiological relevance of the oligomers. In PHb1(*sl*), these residues include Asp42 from  $\beta 2$ , Asp55 and Gln57 from  $\beta 3$ , and Lys63 from  $\beta 4$ , as well as residues from the adjacent protomer: Glu80 from  $\beta 5$  and Asp86, Asp88, and Glu90 from  $\beta 6$  (Fig. 6b). The residues are mostly acidic, except for Lys63 in the middle. Residues from L<sub>12</sub>, L<sub>34</sub>, and L<sub>56</sub> loops contribute to the perimeter of this site (e.g., Ile40, Arg41, Val59, Thr60, and Phe85). The surface is stabilized by residues (Asp40, Arg41, Thr81, Asp86, and Asp88) that form hydrogen bonds with adjacent structural elements. These PHb1 residues are also highly conserved in PHb2. Thus, PHb2 also contains a similar site due to the similarity in the arrangement of its PHb domain in the oligomer (Fig. 6c). The conformational flexibility of L<sub>34</sub> and L<sub>56</sub> loops observed in the crystal structures could be functionally relevant as they are located near the conserved binding sites.

The combination of structural elements from two adjacent PH domains, which form a single binding site, seems to be a novel feature unique to PHb domains, as no similar arrangements are observed in eukaryotic PHs. In fact, the eukaryotic PH domains often exist as a single module in a large protein, and most do not appear to be functionally dependent on oligomerization via their PH domains. One exception is mouse  $\alpha 1$ -syntrophin, which self-associates through its N-terminal PH domain.<sup>28</sup>

The bipolar electrostatic potential distribution (positive and negative potential partitioned between two ends of a molecule) displayed by many PH domains characterized to date is believed to facilitate their interaction with the membrane.<sup>4,6</sup> The PHb domains are mostly electronegative except for a cluster of conserved positively charged residues on the top surface near L<sub>34</sub> (Fig. 6d; Fig. S3). Thus, PHb domains do not display strong bipolarity. The oligomers of PHb1 and PHb2 are also electronegative with electropositive islands located at the perimeter of the dome (Fig. 6e and f). As the conserved binding sites of PHbs consist of mostly acidic residues, decorated with a single basic resi-





**Fig. 6.** Potential binding sites of PHbs. (a) PHb1(sl) pentamer surface colored by sequence conservation (red, conserved; white, non-conserved). (b) Detailed view of the highly conserved residues and surface outlined in (a). (c) Similar conserved site on PHb2. (d–f) Electrostatic potentials of PHb1(sl) monomer, pentamer, and PHb2 dodecamer mapped to their protein surfaces. The color is scaled from  $-5$  to  $5$  kT/e for oligomers and from  $-3$  to  $3$  kT/e for the monomer (blue, positive electrostatic potential; red, negative electrostatic potential). Predicted binding sites are labeled 1–5 and 1–12, and one site is circled on each assembly.

due, electrostatic interaction (charge complementarity) is likely relevant for the function of PHbs.

## Discussion

### Evolution of PH domains

PH-like domains are ubiquitous in eukaryotes. The current SMART database (v6)<sup>29</sup> has cataloged  $\sim 15,000$  proteins (SM00233) with one or more PH domains. Thus, it is surprising that no PH domains of prokaryotic origin have been previously reported. As PH domains are well known to be highly divergent, Ponting *et al.* noted that, if bacterial homologs were to exist, they would likely be undetectable by the sequence methods of the time.<sup>30</sup> With the knowledge that PH-like domains do, indeed, exist in prokaryotes based on the PHb1 and PHb2 structures reported here, we used sequences from the PF08000 family to iteratively identify other potential homologs. Profile-based sequence similarity search methods implemented in PSI-BLAST<sup>31</sup> identified proteins containing a GRAM domain, suggesting a possible evolutionary relationship between PHbs and eukaryotic PH domains. Another profile-based fold recognition method, FFAS, suggested that PHb domains were likely related to the GLUE domain

(97% confidence).<sup>32</sup> However, even with improved methods for detecting remote sequence homology in recent years, the probability scores given by these methods were relatively low. Thus, we further examined the conserved structure features in order to more rigorously establish whether any evolutionary relationship exists between PHb domains and eukaryotic PH domains. In the absence of strong sequence similarity, remote relatives of a highly divergent superfamily can be detected by analyzing unique structural motifs that are important for either fold or function.<sup>33</sup>

The PHb secondary-structure elements are highly related to those of eukaryotic PH-like domains, with similar highly conserved buried hydrophobic residues (Fig. 3b). Besides similarities in the overall fold, we identified several specific structural features that are important to PHb domains, which are also preserved in eukaryotic PH domains. The linker between  $\beta 4$  and  $\beta 5$  in PHb domains consists of a one-turn  $3_{10}$  helix followed by a highly conserved isoleucine (consensus sequence  ${}^{70}\text{PYxxI}{}^{74}$  in PHb1, Fig. 2a). The side chains of Ile74 and Tyr71 point towards the buried hydrophobic cavity consisting of Leu51, Phe77, Leu93, and Ile119. Two hydrogen bonds are observed between main-chain atoms (Tyr71O-Ile74N and Pro70O-Ala73N in PHb1) in the  $3_{10}$  helix preceding Ile74 in PHb1. This region is located at a strategic location where the first  $\beta$ -sheet



( $\beta 1$ – $\beta 4$ ), second  $\beta$ -sheet ( $\beta 5$ – $\beta 7$ ), and the C-terminal helix  $\alpha 1$  pack together and is likely to be important for correct folding of the domain. This structural feature is widely conserved in the PH-like superfamily [e.g., GRAM-PH (PDB ID: 1zvr), GLUE domain (PDB ID: 2cay), and Bruton's tyrosine kinase (PDB ID: 1btk)], supporting its importance for the PH-like fold. Another conserved structural feature of PHb1 domains is a highly conserved, mostly buried salt bridge (Glu31-Arg50 in PHb1, Fig. 3a). Glu31 is located on the loop before  $\beta 1$  and Arg50 at the beginning of  $\beta 3$ . Additionally, Arg50 also connects to  $\beta 4$  and the first  $3_{10}$  helix through hydrogen bonds to Ser68O $\gamma$  and Ile26O. As a result, these interactions link multiple structural elements and likely contribute to structural stability. This ion pair is conserved in eukaryotic PH domains that are the closest structural homologs of PHbs, such as the GRAM-PH domain, GLUE domains, and TFIIF (Fig. 3b). As these key structural features are less likely to have been independently evolved, we suggest that bacterial and eukaryotic PH domains have diverged from a common ancestral fold. The fact that these domains use the same strategy in mediating protein–protein interactions also provides further clues for an evolutionary relationship.

The sequence diversity of PH-like domains could reflect their functional diversity. The binding sites of PH domains are generally functionally dependent and not restricted to a single location.<sup>10</sup> However, with the expansion of our knowledge of PH-like domains, a few functional “hot spots” begin to emerge. One such site is the extended  $\beta 1$ – $\beta 2$  hairpin, which often is required for binding phospholipids. As the  $\beta$ -hairpin in PHb domains is shorter, we expect that they do not utilize this site for functional purposes. Interestingly, eukaryotic PH domains all have longer curved hairpins, even in proteins that do not appear to utilize that site. Thus, the  $\beta 1$ – $\beta 2$  hairpin extension is likely a eukaryotic-specific, structural adaptation that provides new functionality. Further evidence of structural adaptation of the  $\beta 1$ – $\beta 2$  hairpin is provided by PTB domains that contain an additional helix insertion. A second hot spot in PH-like domains is the open edge of  $\beta 5$ , which is frequently used for mediating protein interaction, and it is this site that PHb domains utilize for oligomerization.

#### Unifying uncharacterized protein families: DUF1696 (PF08000), DUF1200, DUF304, and DUF2244

To see whether we could expand the assignment of PHb domains to other families, we looked for relationships of protein families in PFAM using the SCOOP software, which uses a profile-based method to detect distant relationships between protein families.<sup>34</sup> The highest-scoring family to DUF1696 is DUF304 with a score of 15.7. Although this score would normally not be considered significant, it was notable that the second best match was to the GRAM domain (score, 7.9) and the third best match

was to Vps36\_ESCRT-II (score, 6.9), whose structural similarities to PHbs are now confirmed from this work. SCOOP also identifies DUF304 as being related to DUF1200 with a highly significant score of 82.2. DUF2244 was the second highest-scoring match (score, 17.3) and DUF1696 was the third highest (score, 15.7). These analyses suggest that DUF2244 also possibly belongs to the PHb domain family. Based on the known PHb structures, we noted that the current Pfam definition of DUF2244 includes two N-terminal transmembrane helices that anchor this family to the membrane, much like DUF1200. Interestingly, alignments of each of these families suggest that they lack the  $\beta 1$  strand, similar to PA2021, and raise the interesting possibility that PHb domains lacking a  $\beta 1$  strand could represent a more simplified, ancestral-like structure (Fig. S4).

Apart from DUF1696, each of the DUF families that we identify as containing a PHb domain are potential transmembrane proteins where the PHb domain would be located intracellularly as predicted with the Phobius software.<sup>35</sup> The challenge in solving the structure of membrane proteins is well known and perhaps explains why no PHb domains have been characterized previously. Given the potential for PHb domains to form oligomeric rings, we hypothesize that these proteins may act as membrane-bound transporters or pores with the PHb domains mediating oligomerization.

#### Functions of PHb domains in the PF08000 family

As discussed above, a conserved hydrophilic site near the rim of the dome is likely to be of functional importance for both PHb1 and PHb2. The primary difference in their assembly is the size and symmetry of the ring structure; nevertheless, we expect that PHb1 and PHb2 may have similar functions at the molecular level. The physiological functions of PHb domains are currently not understood. Genome-wide studies have shown that the three paralogs of PHb in *B. subtilis* (YjqA, YozO, and YvbH) are expressed, indicating that PHbs are functional in bacteria.<sup>36–38</sup> YozO of *B. subtilis* is induced under antibiotics, heat shock, and alkaline shock, suggesting a role in cell stress responses.<sup>39–42</sup> YozO is under control of sigma factor  $\sigma^W$ , a member of extra-cytoplasmic function (ECF)  $\sigma$  factors that often control functions associated with the cell surface or transport.  $\sigma^W$  likely plays a role in defending the cell against antimicrobial agents.<sup>43</sup> The involvement of the YozO homolog in cell envelope stress response is also confirmed in *Bacillus licheniformis*.<sup>44</sup> Another possible clue for PHb function may be provided by the presence of PHb1 in the lysogenic modules of bacteriophages, such as *Lactococcus* bacteriophage ul36 (ORF124)<sup>45</sup> and Min1 phage from the nematode pathogen *Microbacterium nematophilum* (ORF77).<sup>46</sup> Therefore, PHbs may also play a role in the phage life cycle.

YvbH of *B. subtilis* is found in the sub-membrane fraction and was predicted as a peripheral protein.<sup>38</sup>

PHb1 homologs were also identified in the insoluble sub-proteome of other bacteria, such as *Streptomyces coelicolor*<sup>47</sup> (SCO3793, Swiss-Prot: Q9F325) and *Oceanobacillus iheyensis*<sup>48</sup> (Swiss-Prot: Q8ELK9), an alkaliphilic and halotolerant deep-sea bacterium. These results raise the possibility that the activity of PHb domains may be associated with the membrane. However, the structures of PHbs do not seem to support a direct interaction, since the electronegative conserved binding sites of the PHb oligomers are not suitable for interacting with a negatively charged bacterial membrane. Furthermore, the PHb domain also lacks the actual site and electropositive surface corresponding to the eukaryotic lipid binding site. Instead, we predict that the PHbs likely interact with other proteins through the conserved binding site at their base. A eukaryotic example, which utilizes a similar location for protein interaction, is the TFIIE/TFIIH complex where the  $\beta 5$ – $\beta 6$  hairpin region of TFIIH is involved in the interaction with an extended peptide from TFIIE (Fig. 5). From the molecular architecture point of view, the PHb1 and PHb2 assemblies are similar to “caps” or “domes”. Thus, it is tempting to speculate that the PHb domain is likely an end piece of a molecular assembly, such as a membrane-associated protein complex (where different sizes of PHb domains were evolved to complement different sizes of channels/chambers). It may also be significant that these oligomers are assembled from higher-order cyclic symmetries that are rare in proteins analyzed to date,<sup>21</sup> which may well provide some clues as to function. Further experimental studies are clearly needed for a more complete understanding of the biological function of PHb domains and assemblies, but the identification here of PHb domains provides fascinating new insights into their involvement in all kingdoms of life.

## Materials and Methods

### Protein production

The same protocol was used for the cloning, expression, and production of PHb1(*sl*) (Swiss-Prot: A3QB43), PHb1(*sa*) (Swiss-Prot: A1S3D0), and PHb2 (Swiss-Prot: Q41E03). Clones were generated using the Polymerase Incomplete Primer Extension (PIPE) cloning method.<sup>49</sup> The gene encoding the targeted protein was amplified by polymerase chain reaction (PCR) from genomic DNA of the host bacteria (*Shevanelia* sp. PV-4, *S. amazonensis* SB2B, and *Exiguobacterium* sp. 255-15) using PfuTurbo DNA polymerase (Stratagene) and I-PIPE (Insert) primers that included sequences for the predicted 5' and 3' ends. The expression vector pSpeedET, which encodes an amino-terminal tobacco etch virus (TEV) protease-cleavable expression and purification tag (MGSDKIHHHHHHENLYFQG), was PCR amplified with V-PIPE (Vector) primers. V-PIPE and I-PIPE PCR products were mixed to anneal the amplified DNA fragments together. *E. coli* GeneHogs (Invitrogen) competent cells were transformed with the

V-PIPE/I-PIPE mixture and dispensed on selective LB-agar plates. The cloning junctions were confirmed by DNA sequencing. Using the PIPE method, the section of the gene encoding residues Met1–Met8 was deleted for PHb1(*sl*) (Met1–Phe12 was deleted for PHb2). Expression was performed in a selenomethionine-containing medium. At the end of fermentation, lysozyme was added to the culture to a final concentration of 250  $\mu\text{g}/\text{ml}$ , and the cells were harvested and frozen. After one freeze/thaw cycle, the cells were homogenized in lysis buffer [50 mM Hepes, pH 8.0, 50 mM NaCl, 10 mM imidazole, and 1 mM tris(2-carboxyethyl)phosphine–HCl (TCEP)] and the lysate was clarified by centrifugation at 32,500g for 30 min. The soluble fraction was passed over nickel-chelating resin (GE Healthcare) pre-equilibrated with lysis buffer, the resin was washed with wash buffer [50 mM Hepes, pH 8.0, 300 mM NaCl, 40 mM imidazole, 10% (v/v) glycerol, and 1 mM TCEP], and the protein was eluted with elution buffer [20 mM Hepes, pH 8.0, 300 mM imidazole, 10% (v/v) glycerol, and 1 mM TCEP]. The eluate was buffer exchanged with TEV buffer (20 mM Hepes, pH 8.0, 200 mM NaCl, 40 mM imidazole, and 1 mM TCEP) using a PD-10 column (GE Healthcare) and incubated with 1 mg of TEV protease per 15 mg of eluted protein. The protease-treated eluate was passed over nickel-chelating resin (GE Healthcare) pre-equilibrated with Hepes crystallization buffer (20 mM Hepes, pH 8.0, 200 mM NaCl, 40 mM imidazole, and 1 mM TCEP), and the resin was washed with the same buffer. The flow-through and wash fractions were combined and concentrated for crystallization trials by centrifugal ultrafiltration (Millipore). For PHb1(*sa*), lysines were reductively methylated by adding 40  $\mu\text{l}$  0.98 M dimethylaminoborane and 80  $\mu\text{l}$  3.26% by weight formaldehyde, per milliliter of protein, for 2 h in the presence of crystallization buffer at 4 °C.<sup>50</sup> Methylation reagents were subsequently removed using a PD-10 column.

### Crystallization of PHb1(*sl*)

PHb1(*sl*), concentrated to 13.3 mg/ml, was crystallized by mixing 100 nl protein with 100 nl crystallization solution in a sitting drop above a 50- $\mu\text{l}$  reservoir volume using the nanodroplet vapor diffusion method<sup>51</sup> with standard JCSG robotic crystallization protocols.<sup>12</sup> The crystallization reagent consisted of 37% (v/v) 2-methyl-2,4-pentanediol, 0.15 M sodium chloride, and 0.1 M Hepes, pH 6.83. No additional cryoprotectant was added to the rod-shaped crystal (dimensions:  $\sim 200 \mu\text{m} \times 60 \mu\text{m} \times 40 \mu\text{m}$ ) grown at 277 K.

### Crystallization of PHb1(*sa*)

PHb1(*sa*) crystals were obtained in multiple conditions. The crystal used for refinement was obtained using a precipitating reagent consisting of 10% glycerol, 5% PEG 3000, 20% PEG 400, and 0.1 M Hepes, pH 7.3. (4S)-2-Methyl-2,4-pentanediol was added to the crystal as a cryoprotectant to a final concentration of 10% (v/v). The crystallization reagent yielding the crystal used for the second MAD data set consisted of 5% PEG 3000, 22% PEG 400, 10% glycerol, and 0.1 M Hepes, pH 7.5. Glycerol was added to the crystal as a cryoprotectant to a final concentration of 15% (v/v). Both cubic-shaped crystals (dimensions:  $\sim 30 \mu\text{m} \times 20 \mu\text{m} \times 20 \mu\text{m}$ ) were harvested after 50 days at 293 K. The protein concentration was 19 mg/ml.

## Crystallization of PHb2

The crystallization reagent consisted of 10% (w/v) PEG 6000 and 0.1 M *N,N*-bis(2-hydroxyethyl)glycine, pH 9.0. The diamond-shaped crystal (dimensions: ~40  $\mu\text{m}$   $\times$  40  $\mu\text{m}$   $\times$  40  $\mu\text{m}$ ) was grown at 277 K. Ethylene glycol was added to the crystal as a cryoprotectant to a final concentration of 10% (v/v). The protein concentration was 13.6 mg/ml.

## Diffraction screening and oligomeric state determination

In order to identify the crystals with the best possible diffraction, we screened all harvestable crystal hits for diffraction using the Stanford Automated Mounting system<sup>52</sup> at the Stanford Synchrotron Radiation Light-source (SSRL, Menlo Park, CA). The molecular weight and oligomeric states of PHb1(*sl*) and PHb2 were determined using a 1  $\times$  30 cm Superdex 200 column (GE Healthcare) in combination with static light scattering (Wyatt Technology). The mobile phase consisted of 20 mM Tris, pH 8.0, 150 mM NaCl, and 0.02% (w/v) sodium azide.

## Data collection, structure solution, and refinement

MAD data were collected at the Advanced Light Source (ALS, Berkeley, CA) beamline 8.2.2 (*S. loihica* PHb1) and SSRL beamlines 9-2 (*S. amazonensis* PHb1) and 11-1 (PHb2). Data were collected at wavelengths corresponding to the inflection, high-energy remote, and peak of a selenium MAD experiment at 100 K using ADSC Q315 [PHb1(*sl*)] and MarCCD325 [PHb2 and PHb1(*sa*)] detectors. Data processing, structure solution, and refinement were carried out independently for each of the three proteins using the following protocol. The MAD data were integrated and reduced using XDS and then scaled with the program XSCALE.<sup>53</sup> Selenium sites were located with SHELXD.<sup>54</sup> Phase refinement and automatic model building were performed using autoSHARP,<sup>55</sup> RESOLVE,<sup>56</sup> and wARP.<sup>57</sup> For PHb1(*sa*), an additional MAD data set was collected using another crystal after initial higher-resolution SAD data did not provide enough phasing information. Structural solution was possible with combination of both data sets, which improved the accuracy of phases, especially at lower resolution. Model completion and refinement were performed with Coot<sup>58</sup> and REFMAC5<sup>59</sup> of the CCP4 suite.<sup>60</sup> Tight non-crystallographic symmetry restraints were imposed for most regions of PHb2 except for flexible loop regions, while loose or no non-crystallographic symmetry restraints were used for PHb1s. Each monomer was defined as a TLS group. Experimental phases were used as restraints during refinement. Analysis of the stereochemical quality of the model was accomplished using MolProbity.<sup>14</sup> All molecular graphics were prepared with PyMOL (DeLano Scientific) unless specifically stated otherwise. The alignment of multiple PHb domains and the average rmsd value were calculated using MATT.<sup>61</sup> Pairwise structural comparisons in Tables S4–S6 were calculated with LSQKAB<sup>62</sup> using a maximum common set of C $\alpha$  atoms in both structures. Alignments of PHb domains in Table S7 were calculated using TAlign.<sup>63</sup> Sequence alignments were rendered using TEXshade.<sup>64</sup> Electrostatic potentials were calculated using APBS.<sup>65</sup>

Atomic coordinates and experimental structure factors for both PHb1s at 2.0 Å resolution and PHb2 at 2.42 Å resolution have been deposited in the PDB under accession codes 3dcx, 3hsa, and 3b77, respectively.

## Acknowledgements

The project is sponsored by National Institute of General Medical Sciences Protein Structure Initiative (grant number U54 GM074898 to the JCSG). A.B. and R.D.F. are supported by the Wellcome Trust (grant number WT077044/Z/05/Z). Portions of this research were carried out at the SSRL and the ALS. The SSRL is a national user facility operated by Stanford University on behalf of the U.S. Department of Energy (DOE), Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the DOE, Office of Biological and Environmental Research, and by the National Institutes of Health. The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the U.S. DOE under Contract No. DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory. *Exiguobacterium* sp. 255-15 was a gift from Drs. Tamara Cole and Jim Tiedje, Michigan State University, East Lansing, MI, USA. *Shewanella* sp. PV-4 and *S. amazonensis* SB2B were gifts from Dr. Jim Fredrickson, Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA. The content of this work is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences. We thank Dr. Todd O. Yeates for insightful comments.

## Supplementary Data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2009.11.006](https://doi.org/10.1016/j.jmb.2009.11.006)

## References

- Lemmon, M. A. & Keleti, D. (2005). PH domains. In (Cesareni, G., Gimona, M., Sudol, M. & Yaffe, M., eds), pp. 337–363, Wiley-VCH Verlag GmbH & Co., Weinheim, Germany.
- Rebecchi, M. J. & Scarlata, S. (1998). Pleckstrin homology domains: a common fold with diverse functions. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 503–528.
- Yoon, H. S., Hajduk, P. J., Petros, A. M., Olejniczak, E. T., Meadows, R. P. & Fesik, S. W. (1994). Solution structure of a pleckstrin-homology domain. *Nature*, **369**, 672–675.
- Macias, M. J., Musacchio, A., Ponstingl, H., Nilges, M., Saraste, M. & Oschkinat, H. (1994). Structure of the pleckstrin homology domain from beta-spectrin. *Nature*, **369**, 675–677.
- Timm, D., Salim, K., Gout, I., Guruprasad, L., Waterfield, M. & Blundell, T. (1994). Crystal structure of the pleckstrin homology domain from dynamin. *Nat. Struct. Biol.* **1**, 782–788.
- Ferguson, K. M., Lemmon, M. A., Schlessinger, J. & Sigler, P. B. (1994). Crystal structure at 2.2 Å resolution of the pleckstrin homology domain from human dynamin. *Cell*, **79**, 199–209.
- Lemmon, M. A. (2007). Pleckstrin homology (PH) domains and phosphoinositides. *Biochem. Soc. Symp.* **74**, 81–93.



8. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540.
9. Mayor, L. R., Fleming, K. P., Muller, A., Balding, D. J. & Sternberg, M. J. (2004). Clustering of protein domains in the human genome. *J. Mol. Biol.* **340**, 991–1004.
10. Blomberg, N., Baraldi, E., Nilges, M. & Saraste, M. (1999). The PH superfold: a structural scaffold for multiple functions. *Trends Biochem. Sci.* **24**, 441–445.
11. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S. *et al.* (2004). The Pfam protein families database. *Nucleic Acids Res.* **32**, D138–141.
12. Lesley, S. A., Kuhn, P., Godzik, A., Deacon, A. M., Mathews, I., Kreis, A. *et al.* (2002). Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc. Natl. Acad. Sci. USA*, **99**, 11664–11669.
13. Cruickshank, D. W. (1999). Remarks about protein structure precision. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **55**, 583–601.
14. Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2004). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* **32**, W615–619.
15. Holm, L. & Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480.
16. Begley, M. J., Taylor, G. S., Brock, M. A., Ghosh, P., Woods, V. L. & Dixon, J. E. (2006). Molecular basis for substrate recognition by MTMR2, a myotubularin family phosphoinositide phosphatase. *Proc. Natl. Acad. Sci. USA*, **103**, 927–932.
17. Alam, S. L., Langelier, C., Whitby, F. G., Koirala, S., Robinson, H., Hill, C. P. *et al.* (2006). Structural basis for ubiquitin recognition by the human ESCRT-II EAP45 GLUE domain. *Nat. Struct. Mol. Biol.* **13**, 1029–1030.
18. Teo, H., Gill, D. J., Sun, J., Perisic, O., Vepriyev, D. B., Vallis, Y. *et al.* (2006). ESCRT-I core and ESCRT-II GLUE domain structures reveal role for GLUE in linking to ESCRT-I and membranes. *Cell*, **125**, 99–111.
19. She, M., Decker, C. J., Svergun, D. I., Round, A., Chen, N., Muhrad, D. *et al.* (2008). Structural basis of dcp2 recognition and activation by dcp1. *Mol. Cell*, **29**, 337–349.
20. Jin, R., Junutula, J. R., Matern, H. T., Ervin, K. E., Scheller, R. H. & Brunger, A. T. (2005). Exo84 and Sec5 are competitive regulatory Sec6/8 effectors to the RalA GTPase. *EMBO J.* **24**, 2064–2074.
21. Levy, E. D., Boeri Erba, E., Robinson, C. V. & Teichmann, S. A. (2008). Assembly reflects evolution of protein complexes. *Nature*, **453**, 1262–1265.
22. Hunt, J. F., Weaver, A. J., Landry, S. J., Gierasch, L. & Deisenhofer, J. (1996). The crystal structure of the GroES co-chaperonin at 2.8 Å resolution. *Nature*, **379**, 37–45.
23. Uhlik, M. T., Temple, B., Bencharit, S., Kimple, A. J., Siderovski, D. P. & Johnson, G. L. (2005). Structural and evolutionary division of phosphotyrosine binding (PTB) domains. *J. Mol. Biol.* **345**, 1–20.
24. Hamada, K., Shimizu, T., Yonemura, S., Tsukita, S. & Hakoshima, T. (2003). Structural basis of adhesion-molecule recognition by ERM proteins revealed by the crystal structure of the radixin-ICAM-2 complex. *EMBO J.* **22**, 502–514.
25. Mori, T., Kitano, K., Terawaki, S., Maesaki, R., Fukami, Y. & Hakoshima, T. (2008). Structural basis for CD44 recognition by ERM proteins. *J. Biol. Chem.* **283**, 29602–29612.
26. Zhou, M. M., Ravichandran, K. S., Olejniczak, E. F., Petros, A. M., Meadows, R. P., Sattler, M. *et al.* (1995). Structure and ligand recognition of the phosphotyrosine binding domain of Shc. *Nature*, **378**, 584–592.
27. Okuda, M., Tanaka, A., Satoh, M., Mizuta, S., Takazawa, M., Ohkuma, Y. *et al.* (2008). Structural insight into the TFIIE–TFIIH interaction: TFIIE and p53 share the binding region on TFIIH. *EMBO J.* **27**, 1161–1171.
28. Oak, S. A. & Jarrett, H. W. (2000). Oligomerization of mouse  $\alpha$ 1-syntrophin and self-association of its pleckstrin homology domain 1 containing sequences. *Biochemistry (Moscow)*, **39**, 8870–8877.
29. Letunic, I., Doerks, T. & Bork, P. (2009). SMART 6: recent updates and new developments. *Nucleic Acids Res.* **37**, D229–232.
30. Ponting, C. P., Aravind, L., Schultz, J., Bork, P. & Koonin, E. V. (1999). Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. *J. Mol. Biol.* **289**, 729–745.
31. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
32. Jaroszewski, L., Rychlewski, L., Li, Z., Li, W. & Godzik, A. (2005). FFAS03: a server for profile–profile sequence alignments. *Nucleic Acids Res.* **33**, W284–288.
33. Murzin, A. G. & Bateman, A. (1997). Distant homology recognition using structural classification of proteins. *Proteins*, 105–112.
34. Bateman, A. & Finn, R. D. (2007). SCOOP: a simple method for identification of novel protein superfamily relationships. *Bioinformatics*, **23**, 809–814.
35. Kall, L., Krogh, A. & Sonnhammer, E. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* **35**, W429–432.
36. Wolff, S., Otto, A., Albrecht, D., Zeng, J. S., Buttner, K., Gluckmann, M. *et al.* (2006). Gel-free and gel-based proteomics in *Bacillus subtilis*: a comparative study. *Mol. Cell. Proteomics*, **5**, 1183–1192.
37. Tang, M. R., Sternberg, D., Behr, R. K., Sloma, A. & Berka, R. M. (2006). Use of transcriptional profiling and bioinformatics to solve production problems—eliminating red pigment production in a *Bacillus subtilis* strain producing hyaluronic acid. *Ind. Biotechnol.* **2**, 66–74.
38. Bunai, K., Nozaki, M., Hamano, M., Ogane, S., Inoue, T., Nemoto, T. *et al.* (2003). Proteomic analysis of acrylamide gel separated proteins immobilized on polyvinylidene difluoride membranes following proteolytic digestion in the presence of 80% acetonitrile. *Proteomics*, **3**, 1738–1749.
39. Cao, M., Wang, T., Ye, R. & Helmann, J. D. (2002). Antibiotics that inhibit cell wall biosynthesis induce expression of the *Bacillus subtilis* sigma(W) and sigma(M) regulons. *Mol. Microbiol.* **45**, 1267–1276.
40. Petersohn, A., Brigulla, M., Haas, S., Hoheisel, J. D., Volker, U. & Hecker, M. (2001). Global analysis of the general stress response of *Bacillus subtilis*. *J. Bacteriol.* **183**, 5617–5631.
41. Cao, M., Kobel, P. A., Morshedi, M. M., Wu, M. F., Paddon, C. & Helmann, J. D. (2002). Defining the *Bacillus subtilis* sigma(W) regulon: a comparative analysis of promoter consensus search, run-off

- transcription/microarray analysis (ROMA), and transcriptional profiling approaches. *J. Mol. Biol.* **316**, 443–457.
42. Wiegert, T., Homuth, G., Versteeg, S. & Schumann, W. (2001). Alkaline shock induces the *Bacillus subtilis* sigma(W) regulon. *Mol. Microbiol.* **41**, 59–71.
  43. Helmann, J. D. (2002). The extracytoplasmic function (ECF) sigma factors. *Adv. Microb. Physiol.* **46**, 47–110.
  44. Wecke, T., Veith, B., Ehrenreich, A. & Mascher, T. (2006). Cell envelope stress response in *Bacillus licheniformis*: integrating comparative genomics, transcriptional profiling, and regulon mining to decipher a complex regulatory network. *J. Bacteriol.* **188**, 7500–7511.
  45. Labrie, S. & Moineau, S. (2002). Complete genomic sequence of bacteriophage ul36: demonstration of phage heterogeneity within the P335 quasi-species of lactococcal phages. *Virology*, **296**, 308–320.
  46. Akimkina, T., Venien-Bryan, C. & Hodgkin, J. (2007). Isolation, characterization and complete nucleotide sequence of a novel temperate bacteriophage Min1, isolated from the nematode pathogen *Microbacterium nematophilum*. *Res. Microbiol.* **158**, 582–590.
  47. Kim, D. W., Chater, K., Lee, K. J. & Hesketh, A. (2005). Changes in the extracellular proteome caused by the absence of the bldA gene product, a developmentally significant tRNA, reveal a new target for the pleiotropic regulator AdpA in *Streptomyces coelicolor*. *J. Bacteriol.* **187**, 2957–2966.
  48. Graham, R. L., Pollock, C. E., O’Loughlin, S. N., Ternan, N. G., Weatherly, D. B., Tarleton, R. L. *et al.* (2007). Multidimensional analysis of the insoluble sub-proteome of *Oceanobacillus iheyensis* HTE831, an alkaliphilic and halotolerant deep-sea bacterium isolated from the Iheya ridge. *Proteomics*, **7**, 82–91.
  49. Klock, H. E., Koesema, E. J., Knuth, M. W. & Lesley, S. A. (2008). Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. *Proteins*, **71**, 982–994.
  50. Walter, T. S., Meier, C., Assenberg, R., Au, K. F., Ren, J., Verma, A. *et al.* (2006). Lysine methylation as a routine rescue strategy for protein crystallization. *Structure*, **14**, 1617–1622.
  51. Santarsiero, B. D., Yegian, D. T., Lee, C. C., Spraggon, G., Gu, J., Scheibe, D. *et al.* (2002). An approach to rapid protein crystallization using nanodroplets. *J. Appl. Crystallogr.* **35**, 278–281.
  52. Cohen, A. E., Ellis, P. J., Miller, M. D., Deacon, A. M. & Phizackerley, R. P. (2002). An automated system to mount cryo-cooled protein crystals on a synchrotron beamline, using compact samples cassettes and a small-scale robot. *J. Appl. Crystallogr.* **35**, 720–726.
  53. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800.
  54. Schneider, T. R. & Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **58**, 1772–1779.
  55. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **59**, 2023–2030.
  56. Terwilliger, T. C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **55**, 849–861.
  57. Cohen, S. X., Morris, R. J., Fernandez, F. J., Ben Jelloul, M., Kakaris, M., Parthasarathy, V. *et al.* (2004). Towards complete validated models in the next generation of ARP/wARP. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2222–2229.
  58. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
  59. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53**, 240–255.
  60. Collaborative Computational Project, No. 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 760–763.
  61. Menke, M., Berger, B. & Cowen, L. (2008). Matt: local flexibility aids protein multiple structure alignment. *PLoS Comput. Biol.* **4**, e10.
  62. Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffraction, Theor. Gen. Crystallogr.* **32**, 922–923.
  63. Zhang, Y. & Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309.
  64. Beitz, E. (2000). TEXshade: shading and labeling of multiple sequence alignments using LATEX2 $\epsilon$ . *Bioinformatics*, **16**, 135–139.
  65. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.